logo.jpg

**2016.06.29**

**2820150081 - LEESUJIN**

**2820150089 – KWONDAYEON**

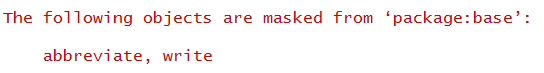
**[ DATAMINING PROJECT – FINAL PROJECT ]**

1. **Installing arules package & loading library(arules)**

**: aruels package is not base package so we need additional installing**

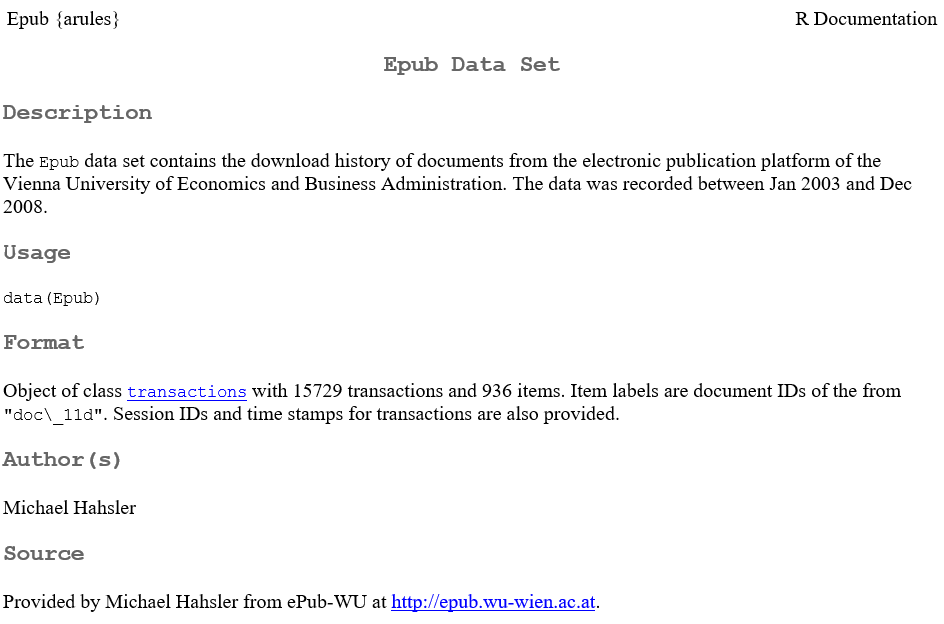
**7.PNG**

**8.PNG**

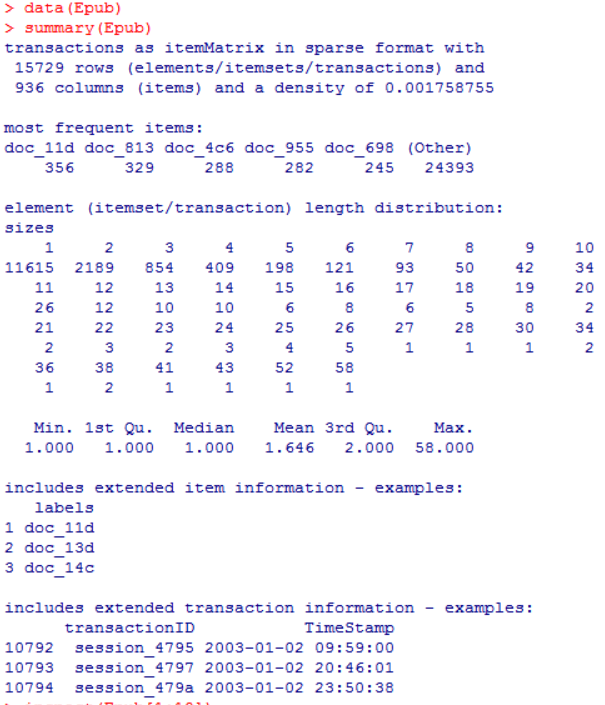
****

1. **Data Secure & exploration**

**: Data used in the analysis is the trading data of Epub data set. Epub data set contains the download history of documents from the electronic publication platform of the Vienna University of Economics and Business Administration. The data was recorded between Jan 2003 and Dec 2008.**



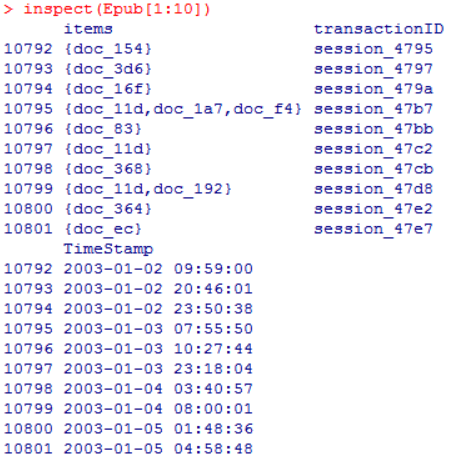
**Loading data(Epub) and summary(Epub) due to view summary information for a data**

****

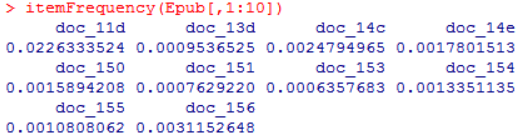
* **ItemMetrix stored in a sparse format transaction data that is**
* **15730 line 936 columns consists**
* **‘most frequent items’ is present most often trading top 5 item name and transaction rate. (doc\_11d is top one)**
* **‘element(itemset/transaction) length distribution’ was indicates whether or nor you had a deal several times by a number of items on one deal in the shopping cart(row=1)**

**We will used inspect() function present the transaction data 10**

**#check itemsets in sparse matrix**

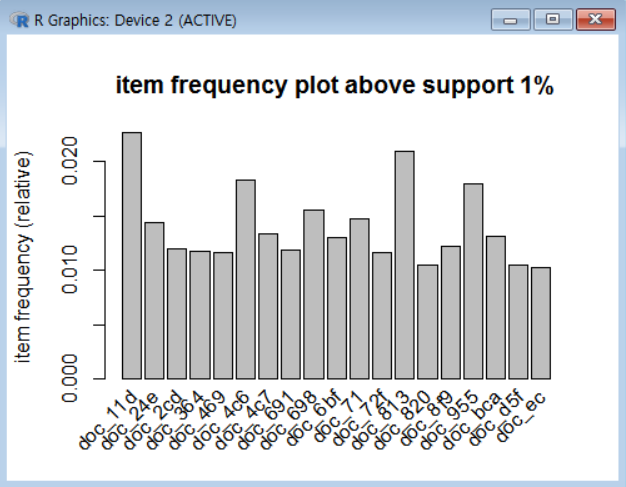
****

**Next we will used itemFrequency() function take a look at the percentage of transactions items per transaction. #support per item: itemFrequency()**

****

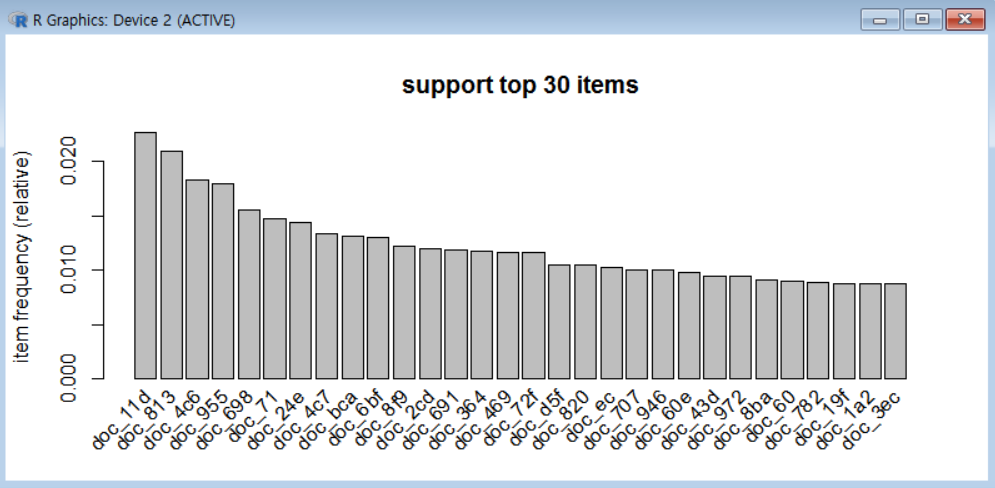
**We used itemFrequencyPlot() function present the item more than 1% of the support and draw bar graph. # item frequency plot : itemFrequentPlot()**

**14.PNG**

****

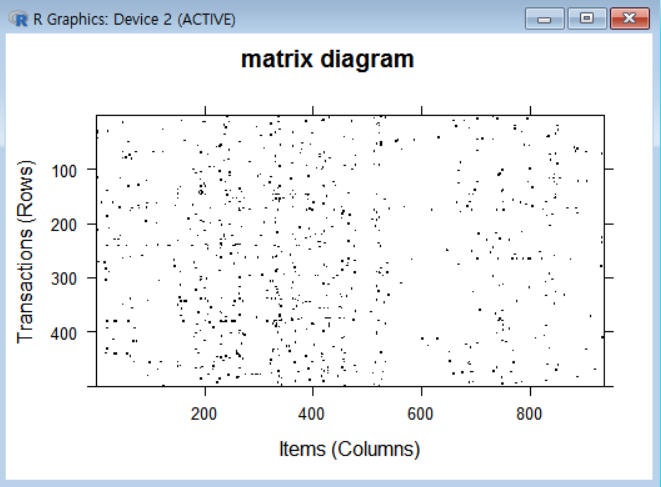
**We used itemFrequencyPloy() function show that support top 30 draw the bar graph # item frequency plot top 30 : itemFrequencyPlot(,topN)**

**1.PNG**

****

**Next is used image() function and sample() function draw diagram random sample of 500 pieces # matrix diagram : image()**

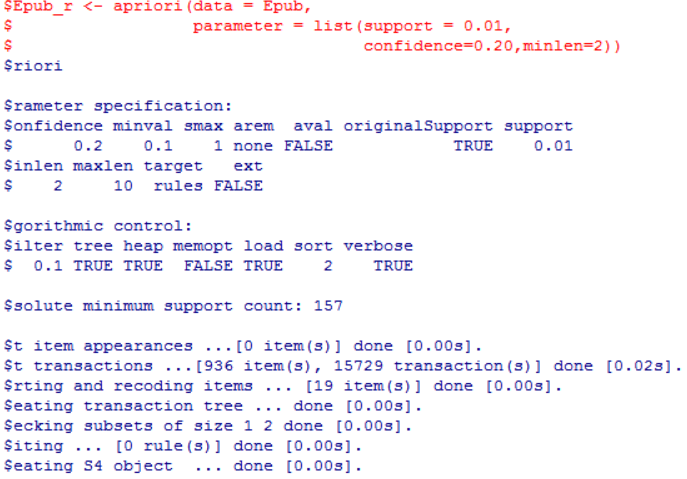
**2.PNG**

****

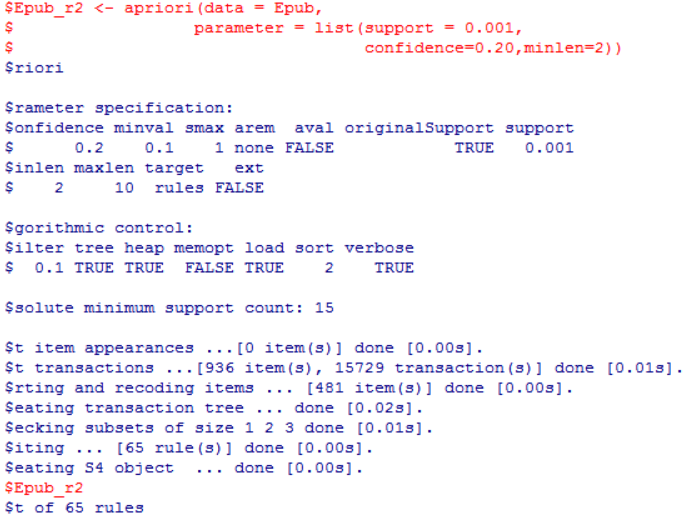
**(in the diagram that plots mean is item transaction occurred)**

1. **Association rule analysis**

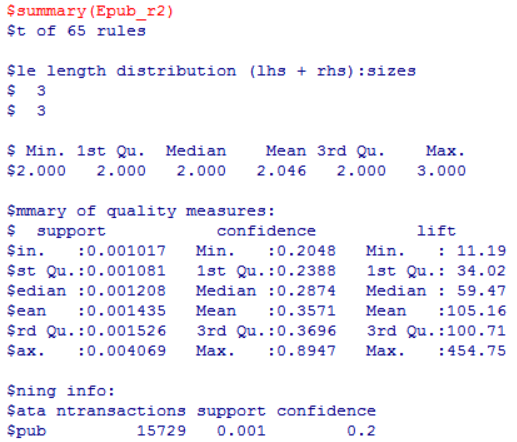
**Now we will analysis using apriori() function of arules package**

****

**But I think first standards too high so association rule appeared ＇０＇，so I tried again by lowering the standards for analysis # re-setting minimum support from 0.01 to 0.001**

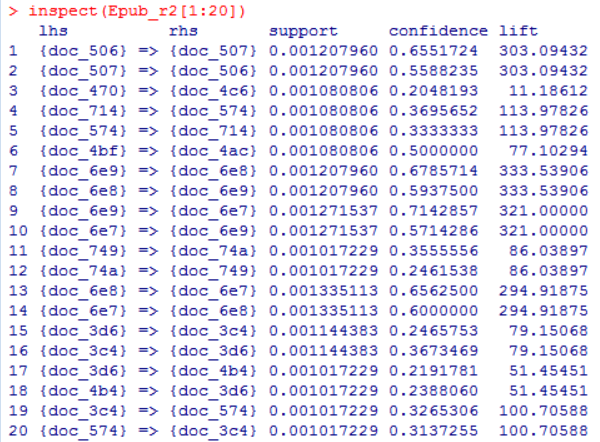
****

1. **Association rules evaluation & inquiry**

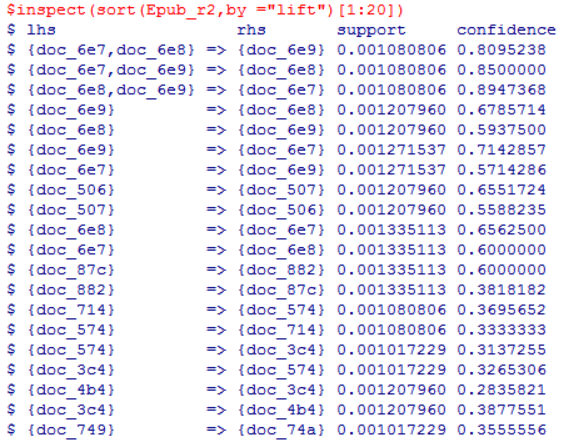
****

**To evaluation the correlation rules to individual rules using the inspect() function**

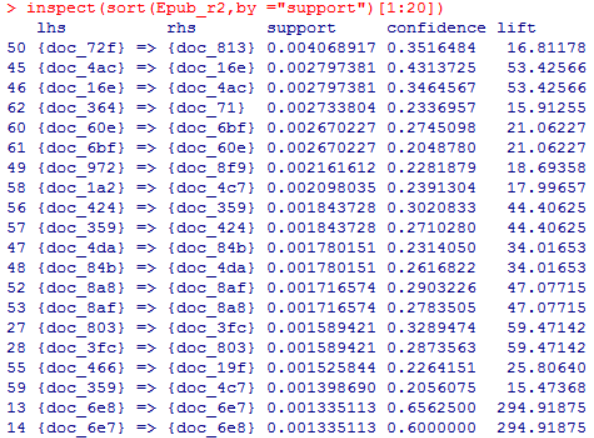
**#inspection of 1~20 association rules : inspect()**

****

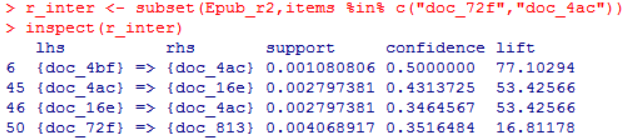
**Align lift the top 20 related rules # sorting association rules by lift : sort( , by=“lift”)**

****

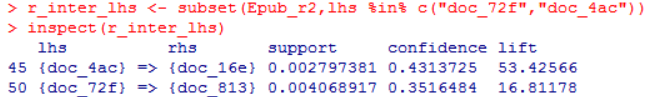
**#sorting association rules by support : sort( , by = “support”)**

****

**How to display the screen that contains the item you are interested in by using the subset function of Association rules. Following is including a using subset() function using association rules “doc\_72f” or “doc\_4ac” on the methods for selecting the rules #subset of association rules : subset()**

****

**If when we find the rule that contains the item when finding association rules to the left of right based on the conditions lhs or rhs like that. # subset with lift-hand side item : subset(lhs %in% “item”)**

****

1. **arulesViz package**

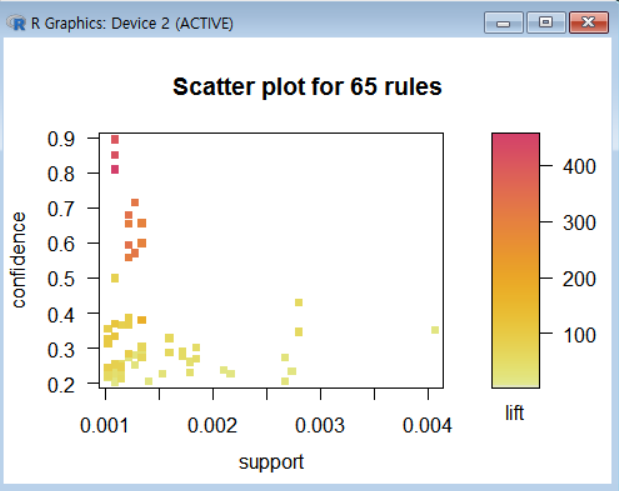
* **Scatter plot for association rules**

**1.PNG**

**2.PNG**

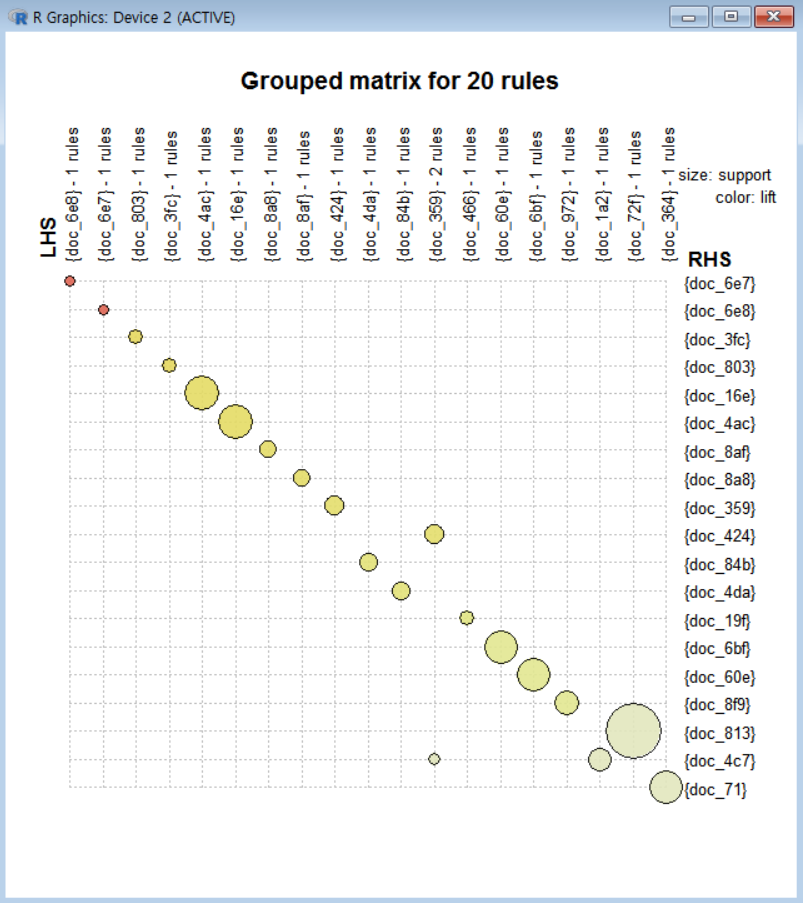
**#scatter plot of association rules**

**3.PNG**

****

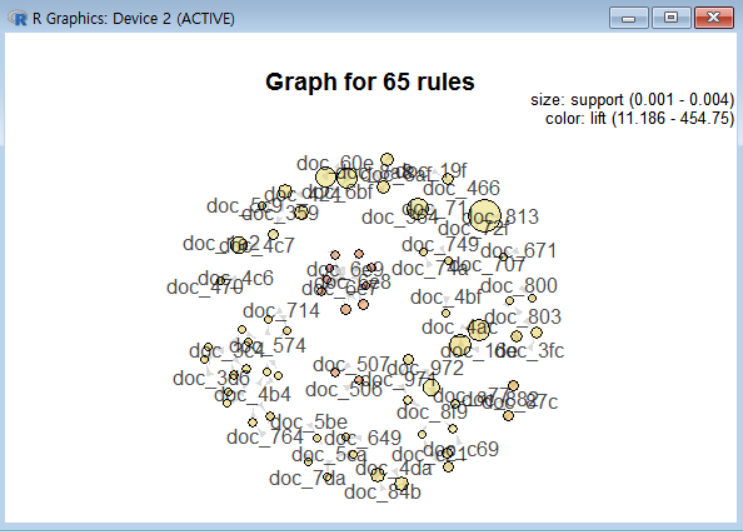
**#Grouped matrix for association rules**

**1.PNG**

****

**#Graph for association rules**

**1.PNG**

****

**Now we Quit method to analysis association rules.**